DOCUMENT RESUME

ED 174 105 HE 011 446

AUTHOR Turner, John D.; Booth, Mary W.
TITLE The Unreliability of Data in the California Community College System. AIR Forum 1979 Paper.
PUB DATE May 79
NOTE 15p.; Paper presented at the Annual Forum of the Association for Institutional Research (19th, San Diego, California, May 13-17, 1979); Appended map may not reproduce well

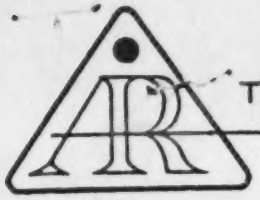
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Centralization; *Community Colleges; Curriculum Evaluation; *Data Collection; Higher Education; *Information Sources; *Institutional Research; Reliability; School Surveys; Sociology; Standards; State Colleges; *Statistical Data; Validity
IDENTIFIERS *AIR Forum 1979; *California Community Colleges; College Catalogs; Grossmont College CA

ABSTRACT

A chronicle of the problems faced in an attempt to collect data on sociology curriculum trends in California's community college system is presented. The project was initiated in an effort to determine if other colleges in the system were experiencing the same difficulties with curriculum and enrollment in sociology courses being encountered by Grossmont College in California. Of the 104 questionnaires sent out to the Deans of Instruction at the existing California community colleges, 84 were completed and returned. Much variety was found in how the data were gathered, and there were difficulties in retrieving other data that were gathered. Often data were unanalyzed or duplicated because of lack of knowledge of what was already stored in the computer. Although college catalogs are the only uniform sources of information available to the public about the curricula at any given college, they were poor sources of accurate information about courses offered. Data published on the system by state and federal agencies were often inaccurate. Several tentative solutions are proposed--such as standards for catalog writing or a centralized data bank--based on the concept that most institutional research is highly unreliable because the data upon which it is based are unreliable. (SC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174105



THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Nineteenth Annual Forum of the Association for Institutional Research held at the Town and Country Inn in San Diego, California, May 13-17, 1979. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC collection of Forum papers.

Mary Corcoran
University of Minnesota
(Editor, AIR Forum Publications)

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NE 011 446

PRE-PUBLICATION COPY. NOT AVAILABLE FOR QUOTATION WITHOUT PERMISSION OF THE AUTHORS.

THE UNRELIABILITY OF DATA IN THE CALIFORNIA COMMUNITY COLLEGE SYSTEM

John D. Turner and Mary W. Booth
Grossmont College
El Cajon, California 92020

Data have become the foundation upon which decision making is built. We have come to depend on them so thoroughly that decisions made in advance of amassing and analyzing data seem to be decisions made with unnecessary risk. Thus, data have become an accepted tool to make objective and scientific a process that had formerly been subjective and unscientific. Data have made large institutions capable of evaluating their various tasks. There is the belief that when the task is sufficiently large to be incomprehensible, data can be rushed in to make the entire ongoing process sensible.

If this process sounds suspiciously like magic, then realize that the institutional researcher has become the Twentieth Century counterpart to the soothsayer of Roman times and the sorcerer of the Dark Ages. But just as soothsayers needed a good set of guts to auger, and just as sorcerers needed the right mysterious ingredients to mix, institutional researchers need good data with which to work. Magic cannot be achieved, and the future cannot be told without the right practitioner mixing the right elements. And just as was true in the less scientific past, finding the best quality ingredients can be a complicated task in itself. We have so many suppliers of data and many of them have credentials that indicate that they should be quality purveyors. The quest for quality data, then, can be as complicated as the formula needed to mix the data into a potent brew.

BACKGROUND OF THE PAPER

Grossmont College is a California community college. In the mid-1970's the college started recording noticeable declines in interest in certain of its Sociology courses. Decreases in certain courses spurred debates within the discipline as to what were the causes and what might be the solutions. This debate raged for several years as decisions were made on which courses and how many sections of various courses should be offered. Suggested solutions differed.

Should new courses be offered to replace those with flagging student interest? Should the curriculum be expanded? Much heat and not much light was generated during these incessant debates. Since everyone seemed to have a pet theory, it was hoped that some statistical data could be found that would provide some insights for solutions. Unfortunately, no such data seemed to be available. There seemed to be only one alternative: seek the data directly and amass as much as possible. Surely, such a collection of data would provide information that would assist in future decision-making.

Thus, in the Fall of 1977, Mary W. Booth and John D. Turner, Sociology instructors at Grossmont College, approached their Dean to seek support for a project that would gather data on Sociology curriculum trends in California's Community College System. We were interested in finding out if other colleges were experiencing the same difficulties with curriculum and enrollment that Grossmont College seemed to be experiencing.

Questionnaires were devised and sent out to the Deans of Instruction at the then existing California Community Colleges (N = 104). The response rate was very good; 84 completed questionnaires were returned. Our task seemed relatively clear-cut: gather the data; code the data; enter the data in a computer; utilize SPSS (the Statistical Package for the Social Sciences) to analyze the data; describe the findings; and write up our conclusions. Easy, yes? No.

From the very outset, it became apparent that there were many anomolies in the data reported. Further, attempts to verify data from other sources only seemed to complicate the confusion. What appeared at first to be a relatively simple task, immediately turned into a mare's-nest.

Thus, this paper is a chronicle of problems encountered in doing a relatively simple piece of institutional research. It is a classic listing of methodological problems in data gathering. It is a look at the magic of institutional research by two experienced practitioners who had been somewhat distant from it for several years. Our experiences with our return to the field may shed some light on the reasons that some of the magic has gone out of the field.

PROBLEMS WITH SOURCES OF DATA

Local Level: Institutional Researchers

When one is shopping for sources of data, the source closest to the site of gathering seems to be the most logical (and easiest) starting point. When one is looking at the California Community College System, the site at which to begin seems to be the individual college. This unit (the college) generates the original

data to be examined. But, since there were 104 such units in California at the time of our survey, there were also 104 distinct administrations, with at least 104 persons involved in institutional research. This allowed for a lot of variety in how data were gathered. In sending questionnaires to these colleges no specific provisions were made to ascertain how the data were gathered.

At some colleges there is a resident institutional researcher hired to perform that job only. At other places a secretary may be responsible for gathering institutional data. Some institutional researchers are employed for gathering data only for state and federal agencies, and may not be utilized to generate in-house research.

In those districts with more than one college, a number of interesting political variations existed. Some multi-campus districts had institutional researchers at the district level who supervised data gathering at all campuses in the district. In other multi-campus districts we were told that each college had its own institutional research expert. One expert commented that this was so each college could compete for funds within the district; it was suggested that the outcome was often doctored data. Thus, it would seem that the position of institutional researcher can be a politically powerful job with opportunities for institutional aggrandizement. If knowledge is power, then data is power's most important handmaiden.

Raw Data:

In our study we found that there were colleges where data was gathered, but retrieval of the data was not easily achieved. Various offices on college campuses have data that they gather at given times during the year, but this data is for use in that office and is often tabulated in such a way that only those in that office can make use of it (i.e., figure out what it means). Typical of such specialized data gathering offices are Registrar's Offices, Offices of Instruction, and Student Personnel Offices.

What occurs when these offices gather data is that they may use it and then allow it to lie unalyzed. Far too often the data needed exists, but no one on the campus can retrieve it or interpret it when retrieved.

College Computers:

College computers are another source for unalyzed data. But here the problems are different. Access to computers is usually established by some pecking order. Knowledge of what is stored in the computer is often governed by a distinctly different pecking order. It is possible that duplication of data takes place because of lack of knowledge about what has already been stored in the computer.

Probably the biggest problem with data in a computer is whether or not it has been verified for accuracy. On a number of occasions we experienced what may have been computer errors generated at other campuses or other agencies. The frightening prospect of an inverted triangle of errors yielding highly skewed data is quite possible with the expanding number of places where the data gets stored. If colleges, districts, the state, and the federal governments each have data stored in computers, the margin of error is compounded greatly. Verification of raw data is a time-consuming and expensive process. It would seem obvious that the person who verifies data would be a likely person to cut in any economy move. In these parlous times of taxpayer revolt, the crucial verification process may become a very low priority item in educational budgets; thus, inaccuracy may be on the increase.

Participation Studies:

In recent years in California, education at the community college level has become a major industry. Competition for funds has become a year-round political combat sport. Since there are so many levels of education in California (K through 12 districts, community college districts, the state college system, and the state university system), private institutional researchers, doing entire system research, in order to give each participating school information, may have become another part of the problem with data reliability.

One such piece of institutional research, the Seamans-Mueller Report, has become the center of much controversy in California. Jean Vincenzi, past President of the Academic Senates of the California Community Colleges has gone on record as stating that the data for Seamans-Mueller is gathered in different ways on different campuses using slightly different formulas. This would obviously skew the results. And since the Seamans-Mueller Report is intended to show weekly Student Contact hours in each college broken into specific comparable categories, this report has become a major political issue in collective bargaining. Whether the various allegations are correct or incorrect, this study demonstrates the difficulties with gathering data independent of government guidelines, when the resulting data are going to be used to influence governmental funding.

If lobbying with data provided by those with a vested interest becomes incorporated into data that is amassed by a governmental agency, there is the potential that the governmental data will reflect the possible bias of the initial gatherer. Without monitoring how the initial study was done, there is no way to ascertain reliability. Data without methodology is difficult to believe.

College Catalogs:

College catalogs are the only uniform source of information available to the public about the curricula at any given college. Yet, as registrars have known for decades, college catalogs can be a very poor source of accurate information about courses offered on any given campus. Often they describe the broadest possible number of concepts that might be touched in a course rather than what is actually taught. These attempts at safe generalities do not make for accurate appraisal. In other cases courses are described that are only irregularly offered. Thus, presumptions that the number of courses mentioned in a catalog are descriptive of the curriculum in any given year are highly suspect. Further, courses sometimes remain in catalogs long after the intention to offer them is past. Catalogs are sometimes mere chronicles of active and dead courses. To distinguish between them is not possible without further investigation.

Merely reviewing course titles does not aid in the attempt to determine parallels between courses offered from one college to another. Courses that are titled similarly may not be parallel in subject matter; courses that are very different in title may be completely comparable in subject matter. Since there is no uniformity of opinion on how cataloging should be done by the faculty of one college, it should be no secret that comparability between colleges is unlikely. Unfortunately, the ability to compare the offerings of one college with another will be, at best, a matter of subjective interpretation -- an art rather than a science.

Typically when registrars confront situations where the catalog description obscures the content of the course, they send for a course syllabus. Should this fail to enlighten, they merely reject the course. Researchers, however, seldom have the time or the staff to be involved in such a process. In most cases they have to determine what criteria will achieve the maximum amount of accuracy within the bounds of their research abilities, time and cost limitations. Catalogs seem to be the best single source of information about curriculum content, but they are fraught with inaccuracies that do not lend themselves to precise quantification. In a situation where the data are so obviously inaccurate one wonders why we continue to act as if catalogs are a source of accurate information.

State Level:

Sometime in the past, the Chancellor's Office divided the California Community Colleges into four regions: North, Bay, Central and South. For historic and political, as well as geographic reasons, we felt it would be extremely useful to analyze our data by regions. However, when we inspected the State's classification system we discovered some illogical anomalies, several of which are cited as follows:

(1) The Peralta Community College District is located in the urban San Francisco Bay area. Feather River College is one of the campuses of the Peralta District. However Feather River is located in Quincy, a town of 2,500 in the middle of the Plumas National Forest, at the junction of state roads 70 and 89, and surrounded by the towns of Massack, Bucks Lake, and Plumas . This single campus is not only non-Bay in character, it is decidedly rural in nature. (See Chart 1)

(2) Geographically Napa and Santa Rosa Colleges are located north of the Bay area, yet they were classified as "Central" by the Chancellor's Office. These colleges are not only north of the San Francisco Bay area, they are actually west of the Central Valley. (See Chart 1)

(3) Although Monterey Peninsula, Gavilan, Cabrillo and Hartnell Colleges are in close geographic proximity (and all south of San Francisco on or near the coast), Monterey and Gavilan had been classified as "South" by the Chancellor's Office, and Gavilan and Cabrillo as "Central." (See Chart 1)

We encountered similar problems with the Chancellor's designation of "Community" as Urban, Suburban or Rural. Whatever the original criteria for classification of colleges/districts, the current scheme is quite questionable (see the "urban" Feather River College example above). Within the San Diego Community College district San Diego City College is definitely located in the urban core, but Mesa College and Mira Mar College are more suburban (some in San Diego would classify Mira Mar as rural); yet the entire District has been classified statistically as Urban by the Chancellor's Office. An updated and revised system is overdue.

Another difficulty with state (and federal) statistical data, relates to the sheer difficulty of compiling huge quantities of data efficiently and accurately, particularly when annual census counts are involved. A case in point involved the "Chancellor's Office California Community Colleges Racial and Ethnic Survey: 1976 District Student Totals." This two-page table included data on total student enrollment, male-female enrollment, and enrollment figures for Blacks, Hispanics, Asians and American Indians. We had hypothesized that interest in particular Sociology offerings could be decidedly affected by such ethnic variables. We proceeded to set up work sheets to analyze state-wide trends by our regional designations.

We were horrified to discover that our totals did not tally with the Chancellor's totals. Booth spent two days trying to locate the error when it occurred to her serendipitously that perhaps the state statistics were in error. By close examination of individual district data, educated hunches, and fortuitous discovery of a single line which had been duplicated for two different districts on each of the two pages of data, Booth located the error--omission of all statistics for San Diego Community College District--Yuba district's data had been duplicated instead. The Chancellor's Office appreciatively responded to the recognition of the error. But the point we wish to make here is that it was more than a year after the data had been collected and published that the error was discovered accidentally and corrected. How many possible secondary errors of analysis may have been generated and perpetuated is beyond both our calculation and comprehension. This, again, points to the need for verification of data.

Federal Level:

A variety of other statistical data is gathered by a host of other agencies. Sometimes the gathered data are primary in nature; sometimes they are merely a secondary compilation from other sources. Various specialized personnel at the district (college) level supply data to various federal agencies. Student Services must supply information on handicapped students to HEW directly, but Affirmative Action officers are directed to compile and self-evaluate data on sex discrimination. This latter data is not sent anywhere but must be kept on hand should HEW personnel drop in to spot-check the data. Affirmative Action officers also must supply annual data for the Federal Equal EmploymentOpportunity Commission (EEOC). As many institutional researchers are well aware, this is only describing the tip of the statistical iceburg.

Filling out the forms for all these various governmental agencies proves confusing, difficult, and time-consuming for individual colleges and college districts. Further, it is confusing, difficult, and time-consuming for various researchers to locate, to dig out, and to evaluate the data which has been generated locally, statewide, and nationally. What we have here is a paper blizzard which is not very selectively evaluated. One often wonders if anyone at any governmental agency actually reads the data--let alone evaluates it. The answer seems to be--seldom. Just keep reporting!! It is not important that there be reliable data from everyone; it is important, however, that everyone have data to report.

PROBLEMS WITH PUBLIC INFORMATION

General:

Perhaps, it is trite to observe that many of us lose sight of the data/forest for the statistics/trees. There are a few researchers who can become so absorbed and challenged by a pile of statistics that they become lost month after month, and year in and year out, and fail to ask some very crucial and fundamental questions about the significance, accuracy and overall reliability of the data they are so busily computing, ponderously analyzing, and pompously expounding upon. Perhaps, it is more than a few of us.

We have had to confront the startling possibility that, perhaps, good institutional research cannot be compatible with good sociological (scientific) research. The possibility cannot escape us that much educational statistical data which generates much political/emotional debate, is analagous to the FBI's Uniform Crime Reports. The crime reports are used by police departments to bolster arguments for increased budgets, increased personnel, or newer hardware. Politicians cite the crime statistics to their own ends. An incumbent district attorney may say "See what a great conviction rate I have; see how crime has held the line." His opponent may counter with, "See how awful things are getting; vote the rascals out and make our community safe for mother, flag, and apple pie." The Uniform Crime Reports are limited in nature. They are compiled at the local level and submitted to the FBI voluntarily. They are only as good or poor as the individually compiled statistics may be (local errors may grossly exaggerate or underestimate the real crime rate)--crimes known to the police are often not reported. Researchers interested in either crime or higher education must often eschew old procedures, techniques, or data for innovative approaches.

State Level:

In California the Chancellor's Office for the Community College system serves as a repository/clearinghouse for much data. Unfortunately, however, the size of the system has led to specialization of functions, sometimes making it difficult to determine which specialized office can be called upon for specific information. While

duplication may result without various agencies being aware of it, it is also possible for the forms sent to individual districts to request the same data using different criteria as well as utilizing different forms.

Much data which is gathered by the Chancellor's Office is geared to statistical summaries compiled for total districts, rather than for individual colleges. As of Fall 1977, there were 70 separate community college districts in the State of California. Of these districts, 53 were single campus districts, and 17 were districts which had two or more individual college campuses (not to be confused with multi-site campuses, where a single campus had satellite clusters in nearby communities). For state administrative purposes this may be exceedingly valuable, but from other perspectives this arrangement of the data leaves much to be desired. For example, the Los Angeles Community College District is composed of nine separate and unique campuses. Los Angeles Pierce College differs widely from Los Angeles Trade Technical College. If one wished to compare enrollment trends in Sociology between suburban, primarily white campuses and urban, high minority enrollment campuses, one would, also, hope the statistics could distinguish between these two Los Angeles campuses. From the data we obtained from the State, this is not possible.

Self-Reported Data:

Almost all levels of governmental data are self-reported rather than gathered by an objective third party. Self-reporting has become the standard manner in which to gather most data. Long ago it was recognized by sociologists studying social stratification and social class that people tend to be more subjective in evaluating themselves than they are in evaluating others. For example, who wants to admit they are lower class when they can say they are really middle class. This tendency to be less than candid in evaluating one's self became known as the "self-aggrandizing effect." One should be skeptical of self-reported data. Many of us find self-aggrandizement useful. Why then should institutions be expected to behave differently from their human components?

Colleges have rules and regulations to determine who can and cannot attend. Students want to attend the college of their choice. When colleges have residency requirements, it is no secret that some students find ways to circumvent the rules. This is merely another form of self-aggrandizement. Students will become whatever the rules/data require them to become.

More capricious, but equally self-aggrandizing, is the occurrence of Anglos listing themselves as Native Americans. Four years ago the Native American Student Association at Grossmont College was astounded at an "official" published figure of the number of Native Americans at the college. Upon investigation, it was found that it had become very fashionable to be identified as Native American. It was not for governmental funds that the students were listing themselves as Native American--it was for social status. Grossmont College, at the time, presumably had a large number of Anglos who were declaring themselves as one-eighth Cherokee. As amusing as this might seem, this self-reported datum greatly angered the members of the Native American Student Association.

In-house competition for funding, classroom space, and advantageous course scheduling has often been fraught with self-aggrandizement; colleges have long been centers for snobbish one-upmanship. If professors are willing to lobby for book shelves in their offices, is it surprising that they would lobby for other perks? The perks of the position may be locally humorous or scandalous, depending on one's sense of humor, but when the lobbying for those perks becomes incorporated in governmental data, the data become suspect. What may be innocent in-house competition may someday be part of an HEW study on classroom utilization. Self-reporting allows and encourages personal and institutional self-aggrandizement.

Colleges that coexist within multiple campus districts are constantly having to justify their requests. Administrators and faculty soon learn that if funding is being determined by the district, it is wise to look more needy than one's sister college. Obviously, this type of competition can lead to self-aggrandizement. Self-reported inaccuracies are compounded by intra-district competition.

As of May 1979, California had108 separate community colleges. The competition for funding from Sacramento and from Washington is tremendous; self-reports and self-aggrandizement abound. The community college system must also compete with the other educational systems in California for funding. With the passage of Proposition 13 in this state, the need to lobby more heavily for funds increases the probability that self-aggrandizement will occur.

In a system where there is so much self-reporting (from the student level all the way to the state level), the opportunity and the necessity for self-aggrandizement increases. One wonders how self-reported data can be believed without some sort of verification process. Even if verification is attempted at each level of this progression, the sanctions against deliberate error are virtually nonexistent. IRS relies on self-reported data, but spot checks by the IRS can result in criminal penalties for

fraudulent data. Erroneous educational data may, at most, result in embarassment to a college.

TENTATIVE SOLUTIONS

Any attempt at a solution to a problem presumes that there is some uniformity of agreement on the problem. Defining the problem with data gathering procedures is determined by how much accuracy is needed for data to be considered reliable. Thus, our tentative conclusions are intended to open debate about how much reliability is necessary in institutional research. Our list of possible solutions is based upon the concept that most institutional research is highly unreliable because the data upon which it is based is unreliable.

1. Most statistics generated for governmental purposes do not indicate the methodology used in the compilation and analysis. Public information about methods used in the gathering are as important as, if not more important than, the final results.

2. Many colleges and universities have people highly trained in data gathering who are not part of the institutional research office. They seem to be underutilized. Perhaps, a pool of experts on each college campus could be useful in developing and refining data gathering practices at the local and other levels.

3. If college catalogs are to continue to be used as a source of data, perhaps some new standards for writing catalogs ought to be mandated. If this is too politically risky, maybe a new system that simply reflected what is actually being offered could be developed and stored separately for use by researchers.

4. Who defines a problem determines how it will be solved; therefore, an open dialogue on the problems of data gathering ought to be an on-going process. It ought to employ the broadest possible cross-section of experts from all levels. Institutional researchers defining the problem may not always unearth the vastness of a given problem.

5. Once data are gathered, they ought to be readily available to others. Ways must be found to permit the greatest possible access to these data. Data gathered but unavailable to others is a waste of valuable resources.

6. A central data bank with a descriptive bibliography might be made available at the regional level. Data could be stored on how to retrieve specific information.

7. Primary data should always be a researcher's first choice. Data banks of verified primary data need to be made available to all researchers by governmental agencies. Relying on secondary data is always risky.

8. An index of reported data held by various agencies and/or subdivisions of agencies may prevent various governmental fiefdoms from duplicating efforts.

9. Studies that are done by outside consultants ought to be standardized by governmental agencies that utilize such sources. Strict standards of methodology should be demanded by all agencies that use independent data.

10. Standardization of data gathering procedures ought to be maintained by all levels of government when they are compiling primary data.

11. The problems inherent in self-reported data ought to be thoroughly acknowledged. Perhaps a period of amnesty for our present data gathering procedures could be declared. Rewards could be conferred on researchers who listed as many reliability problems as they knew along with tentative solutions to the problems. Idealistic as they may seem there is proven merit in having police learn from burglars. In like kind, it may be instructive to learn about self-aggrandizement from experts. The nature of the system that fosters it ought to be investigated as well.

12. With most of the present data gathering procedures it is probably impossible to predict the degree of error built into any data collected by a governmental agency. If we do not know how unreliable our data are, how can we use them to be predictive?

13. All research gathering organizations should be encouraged to reevaluate specific classification criteria (demographic, income, etc.) on a regular on-going basis to avoid obvious illogical and anamalous biases in our procedures.

14. Data ought not to be exclusively self-reported unless we can devise methods of verification that are effective. Either a system of incentives and/or a system of penalties ought to be established to bring this about.

CONCLUSION

Once data were thought of as the magic that worked. They do not seem to be working any more. Perhaps it is the quality of the data that is the main problem . . . . But then again, perhaps data never was magic! Such a conclusion may seem naive. So did the innocent observation, "But the Emperor has no clothes!"

CHANCELLOR'S OFFICE GEOGRAPHIC LOCATIONS

of the California Community Colleges